

Weekly Report

2017.06.25 刘良军

本周工作

1. 找到一个高维的数据集，是用于表示在 twitter 上使用的字符标点的特征的，对这个数据集进行处理，使其能用于我们的系统。
2. 针对 Mnist（手写数据）数据集和 Twitter 数据集，用我们的方法和 Largevis 两种方法对其进行高维数据投影到低维（现在是二维），比较其 knn 的准确率，knn 生成时间，以及投影时间。

下周工作

1. 现在的系统读较大的数据集时会占有很大的内存，可能存在一定的内存问题，后面可能要进行优化。
2. 投影时会消耗很长时间，下一步了解投影原理，看看能不能进行优化。
3. 对后面找的 Twitter 的数据集，用我们的方法生成的 knn 速度相对较快，但准确率不稳定（准确度时高时低），可能是计算准确度的方法的问题。
4. 只有手写数据比较适合投影，可能还需要找一些新的适合投影的数据来对系统进行测试。
5. 和朱博多多交流。